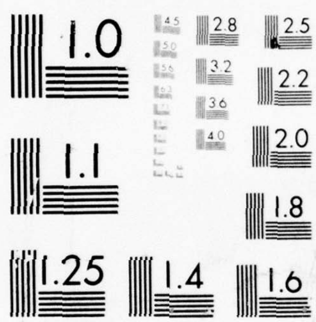
AD-A050 096 NAVAL RESEARCH LAB WASHINGTON D C F/G 9/4
AXIOMATIC CHARACTERIZATION OF A FAMILY OF INFORMATION MEASURES --ETC(U)
NOV 77 R W JOHNSON
UNCLASSIFIED NRL-MR-3646 SBIE-AD-E000 078 NL

| OF |

AD
A050 096

END
DATE
FILMED
3-78
DDC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

1

ade 000078

NRL Memorandum Report 3646

Axiomatic Characterization of a Family of Information Measures that Contains the Directed Divergences

RODNEY W. JOHNSON

Information Systems Staff
Communications Services Division

74

ade 000078

November 1977

NAVAL RESEARCH LABORATORY
Washington, D.C.

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER NRL Memorandum Report 3646	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER 9
4. TITLE (and Subtitle) 6 AXIOMATIC CHARACTERIZATION OF A FAMILY OF INFORMATION MEASURES THAT CONTAINS THE DIRECTED DIVERGENCES.		5. ~~TYPE OF REPORT~~ & PERIOD COVERED Interim report on a continuing ~~NRL problem~~.
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) 10 Rodney W. Johnson		8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS Naval Research Laboratory Washington, D.C. 20375		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NRL Problem B02-35 16 RRØ14-09 17 RRØ14-09-41 PE 61153N
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Washington, D.C.		12. REPORT DATE 11 November 1977
		13. NUMBER OF PAGES 28
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office) 12 29p.		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT (of this Report) Approved for public release; distribution unlimited. 14 NRL-MR-3646		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report) 18 SBIE 19 AD-E000 078		
18. SUPPLEMENTARY NOTES		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) Information measures Directed divergence Information axioms Prior probability		

DDC
RECEIVED
FEB 17 1978
B

the integral

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Let p and q be probability densities. In the terminology of S. Kullback, the directed divergences of p and q are given by $\int p(x) \log (p(x)/q(x))\, dx$ and by the same expression with p and q interchanged; the divergence is the sum of the directed divergences. These quantities have applications in information theory and to the problem of assigning prior probabilities subject to constraints. In this report, it is shown that the directed divergences and their positive linear combinations, including the divergence, are characterized by axioms of positivity, additivity, and finiteness; in the course of the

(Continues) next page

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

251 950

LB

20. *Abstract (Continued)*

proof, the latter two are shown to imply yet another axiom: invariance. These axioms are fundamental in work on prior probabilities. It has been claimed that they characterize only constant multiples of the single directed divergence; that claim is here refuted.

CONTENTS

AXIOMATIC CHARACTERIZATION OF A FAMILY OF INFORMATION MEASURES THAT CONTAINS THE DIRECTED DIVERGENCES

I. INTRODUCTION

R. L. Kashyap, in [1], has considered a system characterized by a vector Y of random outputs and a vector Λ of certain other parameters such that the conditional probability density $p(Y|\Lambda)$ is some known function. The output Y is of direct concern and directly accessible to measurement; Λ is not, except insofar as it affects Y. It is assumed that Λ is distributed with a probability density $f(\Lambda)$, which Kashyap calls the "true" density. Together with the known density $p(Y|\Lambda)$, this determines a probability density p(Y) for Y. Neither $f(\Lambda)$ nor p(Y) is given, although f may be subjected to certain known convex constraints. Without the knowledge that defines the density $f(\Lambda)$ one has the problem of assigning a "prior" density $b(\Lambda)$ that reflects the knowledge one does have: the conditional density $p(Y|\Lambda)$ and the constraints. Additional information, such as particular measurements of Y, may be folded in by Bayes's theorem. But it is the important problem of choosing the prior that concerns Kashyap.

Among previously suggested guides to choosing a prior is the principle of maximum entropy [2]: the "most noncommittal" prior is the one that maximizes entropy subject to the given constraints. Kashyap proposes an alternative, the "principle of min-max uncertainity." He defines an "uncertainty functional" $\varphi(\Lambda;Y)$, an information-theoretic measure of the discrepancy between the probability density for Y that we assign, defined by $b(\Lambda)$ and $p(Y|\Lambda)$, and the "true" density, defined by $f(\Lambda)$ and $p(Y|\Lambda)$. The proposed principle is to choose b so as to place the lowest possible upper bound on the largest value $\varphi(\Lambda;Y)$ can take for any f satisfying the constraints -- that is, choose b to minimize the maximum possible uncertainty. The functional is defined so as to satisfy certain plausible axioms of additivity, coordinate invariance, and the like, imposed on φ and an auxiliary functional ψ; Kashyap claims that the axioms in fact determine φ uniquely, up to a pair of arbitrary constants. That claim is the substance of Theorems 1 and 2 of his paper.

However, the claim is erroneous; the theorems are false. In [3] we point out the errors in the two proofs and exhibit counterexamples: functionals that satisfy the axioms in the paper but form a family larger than the family of functionals presented there. We also correct Kashyap's Lemma 1, which is false as stated but can be fixed up well enough to serve its purpose. Our concern here is to correct Kashyap's Theorem 1 by finding the complete set of functionals determined by the axioms imposed on ψ. We also show that one of the axioms is redundant, being a consequence of the others.

Note: Manuscript submitted October 27, 1977.

Theorem 1 deals with the special case when the "true" distribution of Λ is concentrated at a single point λ^* (so $f(\lambda) = \delta(\lambda - \lambda^*)$) and the uncertainty functional $\varphi(Y;\Lambda)$ reduces to the form $\psi(\lambda^*;q(Y))$, where

$$\psi(\lambda^*;q(Y)) = \int d|y|\, L\,[p(y|\lambda^*),\, q(y)]$$

for some function L. The axioms imposed on ψ are:

<u>Positivity</u>: $\psi(\lambda^*;q(Y)) \geq 0$ with equality only if $q(Y) = p(Y|\lambda^*)$.

<u>Invariance</u>: The uncertainty is invariant under nonsingular linear transformations on Λ and Y.

<u>Additivity</u>: The uncertainty functional for a composite of two independent systems is the sum of the uncertainty functionals for the two component systems.

(For less abbreviated statements, see [1].)

Theorem 1 asserts that such a ψ can only be of the form

$$\psi(\lambda^*,q(Y)) = C_3 \int d|y|\, p(y|\lambda^*) \log \frac{p(y|\lambda^*)}{q(y)}$$

for some constant $C_3 > 0$. But actually the two-parameter family

$$\psi(\lambda^*;\, q(Y)) = C_2 \int dy\, q(y) \log \frac{q(y)}{p(y|\lambda^*)} + C_3 \int dy\, p(y|\lambda^*) \log \frac{p(y|\lambda^*)}{q(y)} \qquad (1)$$

satisfies the axioms, including positivity if C_2 and C_3 are nonnegative and not both zero. Kashyap obtains only the second term. When $C_2 = C_3 = 1$, Kullback [4, pp. 6,7] calls the two terms "directed divergences" and their sum the "divergence."

Kashyap's Theorem 2 deals with the general case, when the "true" distribution of Λ need not be concentrated at a single point. The uncertainty functional is assumed to have the form

$$\varphi(\Lambda;Y) = \int d|y|\,d|\lambda|\, L\,[p(y|\lambda),\, p(y),\, q(y),\, f(\lambda),\, b(\lambda)]$$

for some function L (different from the previous L). The axioms imposed are:

Positivity: $\varphi(\wedge;Y) \geq \int d|\lambda|\ f(\lambda)\psi[\lambda;q(Y)] \geq 0.$

Consistency with ψ: If $f(\lambda) = \delta(\lambda-\lambda^*)$, then $\varphi(\wedge;Y) = \psi(\lambda^*;q(Y))$.

Invariance: (As for ψ above).

Additivity: (as for ψ above).

Theorem 2 asserts that such a φ can only be of the form:

$$\begin{aligned}\varphi(\wedge;Y) &= C_1\zeta[f(\wedge)] + C_3\int d|\lambda|\ f(\lambda)\ \psi(\lambda;q)\\ &= C_1\zeta[f(\wedge)]\\ &\quad + C_3\int d|\lambda|\ d|y|\ f(\lambda)p(y|\lambda)\log\frac{p(y|\lambda)}{q(y)},\end{aligned}$$

where ζ is the mutual information:

$$\zeta[f(\wedge)] = \int d|\lambda| d|y| f(\lambda)p(y|\lambda)\log\frac{p(y|\lambda)}{p(y)},$$

and C_3 is as in the previous discussion[1] of ψ.

However, the axioms are satisfied by at least the following 5-parameter family of functionals:

$$\begin{aligned}\varphi(\wedge;Y) = C_1&\int d\lambda\ dy\ f(\lambda)p(y|\lambda)\log\frac{p(y|\lambda)}{p(y)}\\ &+ C_2\int d\lambda\ dy\ f(\lambda)q(y)\log\frac{q(y)}{p(y|\lambda)}\\ &+ C_3\int d\lambda\ dy\ f(\lambda)p(y|\lambda)\log\frac{p(y|\lambda)}{q(y)}\\ &+ C_4\int d\lambda\ dy\ f(\lambda)p(y)\log\frac{p(y)}{p(y|\lambda)}\\ &+ C_5\int d\lambda\ dy\ f(\lambda)q(y)\log\frac{p(y)}{p(y|\lambda)}\end{aligned}$$

---

1 We retain the explicit constant where Kashyap elects to set $C_3 = 1$ by convention.

with all $C_i \geq 0$. Kashyap obtains only the first and third terms.

Our counterexamples to Theorems 1 and 2 leave open the question whether there are still further functionals satisfying the axioms. We will argue in Section III that the two-parameter family (1) is essentially the complete solution for ψ. We make no claims of completeness for the five-parameter family of functionals φ. The theorem in Section III actually states that ψ is characterized by the positivity and addivity axioms alone. The invariance axiom is not needed as a hypothesis since it is a consequence of positivity and additivity, as we show in Section II.

The characterization of ψ is similar to the characterization of the directed divergence given by Pl. Kannappan [5,6]. Kannappan's result is stated for finite, discrete distributions $p = (p_1, \ldots, p_n)$ and $q = (q_1, \ldots, q_n)$, where q, but not p, is allowed to be incomplete: that is, $\Sigma_i \; q_i \leq 1$, but equality is not demanded. For a discussion and other references, see the book by Aczél and Daróczy [7].

II. DERIVATION OF INVARIANCE FROM ADDITIVITY

At this point we abandon Kashyap's notation. The functional F and function f in what follows correspond to his ψ and L.

In this section we prove a theorem to the effect that any functional satisfying the additivity requirements imposed on ψ (= F) must also satisfy the invariance requirement, even if we restrict attention to probability densities p and q that nowhere take the value zero. The restriction sidesteps problems with division by zero which would otherwise arise, since (1) involves p/q and q/p. With general densities p and q, we could adopt some convention like setting $0 \log 0/0 = 0$, but we defer consideration of such matters to Section IV. We will also see in Section IV that, even with the restriction to positive probability densities p and q, it is impossible to avoid problems with divergent integrals that may lead to infinite values of F. For the present we will therefore adopt, where appropriate, the additional hypothesis $F(p,q) < \infty$, meaning that the integral defining F(p,q) converges absolutely to a finite value. We also add a mild finiteness requirement, $F(p,p) < \infty$, to the axioms. It then becomes possible to prove that in fact $F(p,p) = 0$. We can even prove that $f(t,t) = 0$ for all $t > 0$; this is one of the lemmas used in the proof of the theorem.

<u>Theorem A.</u>

Let f be a function[2] of two real variables, and define a functional F by setting

$$F(p,q) = \int f(p(x), q(x))\, dx$$

whenever p and q are positive probability-density functions on a linear space[2]. Suppose F satisfies the following two axioms.

<u>Finiteness</u>. $F(p,p) < \infty$.

<u>Additivity</u>. If the space on which p and q are defined is the product of two linear spaces X' and X", and p and q have the product form

$$p(x', x'') = p'(x')p''(x'')$$

$$q(x', x'') = q'(x')q''(x'')$$

in terms of probability-density functions p', q' on X' and p", q" on X", then

$$F(p,q) = F(p',q') + F(p'',q'').$$

---

[2] We write <u>function</u> for <u>real</u>, <u>measurable</u> <u>function</u>; <u>linear</u> <u>space</u> for <u>finite-dimensional</u>, <u>real</u> <u>linear</u> <u>space</u>; and <u>set</u> for <u>measurable</u> <u>set</u>.

Then F also satisfies the following axiom, at least if $F(p,q) < \infty$ or $F(p',q') < \infty$.

Invariance. If T is a nonsingular linear transformation on the domain of p and q and we define probability densities p' and q' by

$$p'(x) = p(Tx) \det T,$$

$$q'(x) = q(Tx) \det T,$$

then

$$F(p',q') = F(p,q).$$

Lemma A.1.

Assume the hypotheses of Theorem A. Let p' and q' be positive probability densities with $F(p',q') < \infty$. Then

$$\int f(p'(x), q'(x))\, dx - (1/s) \int f(sp'(x), sq'(x))\, dx = f(1,1) - f(s,s)/s$$

for every real number $s > 0$.

Lemma A.2

Assume the hypotheses of Theorem A. Then

$$f(t,t) = 0$$

for every real number $t > 0$.

Proof of Lemmas.

To get information about $f(t,t)$ for some particular real numbers t, we will apply the additivity axiom to density functions that assume constant values on certain intervals. Consider any two numbers $u > 0$ and $s > 0$.

Define $p_1''(x'') = q_1''(x'') = u$ when $0 \le x'' \le a$, where a is chosen so that $au < 1$. Off the interval $A = [0,a]$, define p_1'' in any way that makes p_1'' a positive probability-density function on the real numbers; let $q_1'' = p_1''$. Define

$$p_2''(x'') = q_2''(x'') = \begin{cases} p_1''(x'') & x'' < 0 \\ su, & 0 \le x'' \le a/s \\ p_1''(a + x'' - a/s), & a/s < x'' \end{cases}$$

(see Figure 1). Then $p_2'' = q_2''$ is a positive probability density. By the additivity axiom,

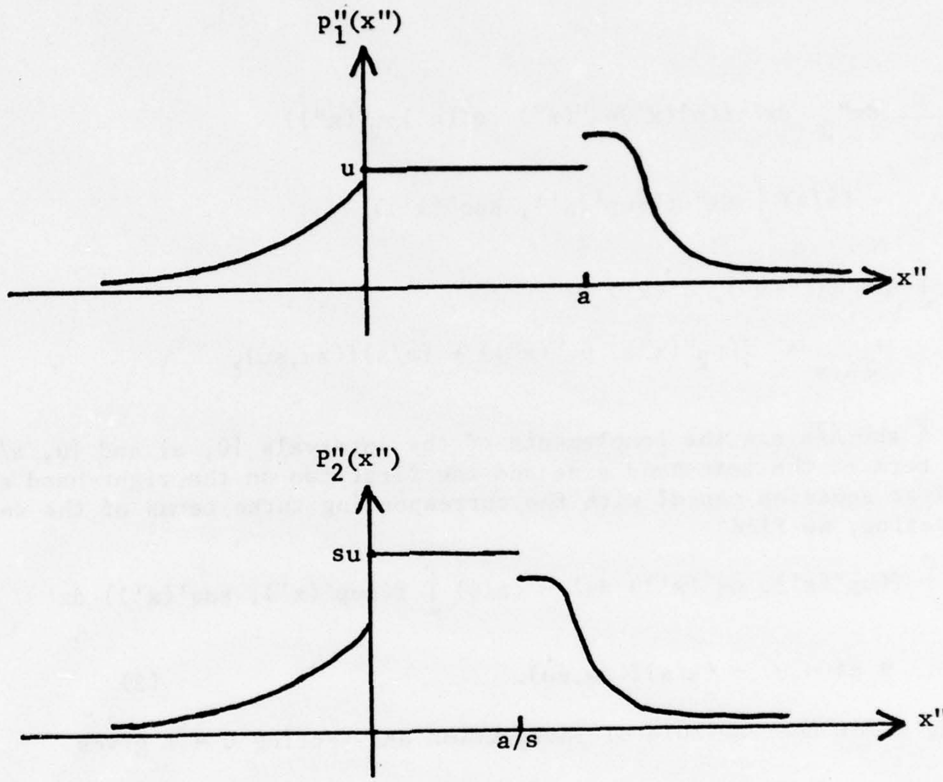


Figure 1

$$\int_{\bar{A}} dx'' \int dx' \; f(p'(x')p\,''(x''),\; q'(x')p_1''(x''))$$

$$+ \; a \int dx' \; f(up'(x'),\; uq'(x'))$$

$$= \int dx' \; f(p'(x'),\; q'(x'))$$

$$+ \int_{\bar{A}} dx'' \; f(p_1''(x''),\; p_1''(x'')) + af(u,u)$$

and

$$\int_{\overline{A/s}} dx'' \int dx' \; f(p'(x')p_2''(x''),\; q'(x')p_2''(x''))$$

$$+ \; (a/s) \int dx'' \; f(sup'(x'),\; suq'(x'))$$

$$= \int dx' \; f(p'(x'),\; q'(x'))$$

$$+ \int_{\overline{A/s}} dx'' \; f(p_2''(x''),\; p_2''(x'')) + (a/s)f(su,su),$$

where $\bar{A}$ and $\overline{A/s}$ are the complements of the intervals $[0, a]$ and $[0, a/s]$. The first term on the left-hand side and the first two on the right-hand side of the first equation cancel with the corresponding three terms of the second. Subtracting, we find

$$a \int f(up'(x'),\; uq'(x'))\, dx' - (a/s) \int f(sup'(x'),\; suq'(x'))\, dx'$$

$$= af(u,u) - (a/s)f(su,su). \qquad (2)$$

Writing x for the variable of integration and setting $u = 1$ gives

$$\int f(p'(x),\; q'(x))\, dx - (1/s) \int f(sp'(x),\; sq'(x))\, dx$$

$$= f(1,1) - f(s,s)/s \quad . \qquad (3)$$

This establishes Lemma A.1.

Now take any probability densities p,q,p',q' related by

$$p(x) = up'(ux),$$

$$q(x) = uq'(ux),$$

and such that $F(p,q) < \infty$, $F(p',q') < \infty$. By Lemma A.1, equation (3) still holds when we replace p' and q' by p and q, since p and q satisfy the hypotheses as well as p' and q'. Therefore, we may substitute $up'(ux)$ and $uq'(ux)$ for $p'(x)$ and $q'(x)$ in (3). We get

$$\int f(up'(ux),\ uq'(ux))\ dx - (1/s) \int f(sup'(ux),\ suq'(ux))\ dx$$
$$= f(1,1) - f(s,s)/s.$$

But a change of variables $x' = ux$ in (2) gives

$$u \int f(up'(ux),\ uq'(ux))\ dx - (u/s) \int f(sup'(ux), suq'(ux))\ dx$$
$$= f(u,u) - f(su,su)/s \quad .$$

By the last two equations,

$$f(1,1) - f(s,s)/s = f(u,u)/u - f(su,su)/su.$$

Define the function h by

$$h(t) = f(t,t)/t - f(1,1) \ ;$$

then

$$h(su) = h(s) + h(u) \quad .$$

The solutions of this equation, aside from non-measurable functions, are all of the form

$$h(t) = a \log t$$

for some constant a; thus

$$f(t,t) = at \log t + bt,$$

where $b = f(1,1)$. Thus,

$$F(p,p) = a\int p(x) \log p(x)\ dx + b\int p(x)\ dx.$$

If $a \neq 0$, we can violate the finiteness axiom by taking a positive probability density p for which the first integral diverges; therefore, $a = 0$, and $F(p,p) = b$. Now the additivity axiom implies that $b = b + b$; therefore, $b = 0$. Thus,

$$f(t,t) = 0.$$

This establishes Lemma A.2.

<u>Proof</u> <u>of</u> <u>Theorem</u>.

For any positive probability densities p and q and any nonsingular linear transformation T (all on the same linear space), define

$$p'(x) = p(Tx) \det T$$

$$q'(x) = q(Tx) \det T,$$

and set $s = 1/\det T$. We may assume $F(p',q') < \infty$; if not, exchange the roles of the primed and the unprimed quantities and replace T by its inverse. By the two lemmas,

$$\int f(p'(x),\ q'(x))\ dx = (1/s) \int f(sp'(x),\ sq'(x))\ dx.$$

But then

$$\int f(p'(x),\ q'(x))\ dx = \int f(p(Tx),\ q(Tx))\ (\det T)\ dx.$$

The left-hand side is $F(p',q')$. A change of variables makes the right-hand side $F(p,q)$. Therefore the invariance axiom holds.

III. THE MAIN THEOREM

In this section we prove a theorem that characterizes all the functionals F that satisfy the positivity and additivity axioms (together with finiteness of $F(p,p)$). As a corollary, we obtain a similar theorem in which the positivity condition $F(p,q) \geq 0$ is replaced by a semiboundedness assumption $F(p,q) \geq F(p,p)$.

Theorem B.

Let f be a function of two real variables and define a functional F by setting

$$F(p,q) = \int f(p(x),q(x))dx$$

whenever p and q are positive probability-density functions on a linear space. Suppose F satisfies the finiteness and additivity axioms of Theorem A together with the following axiom.

Positivity. $F(p,q) \geq 0$ with equality only if $p = q$.

Then F has the form

$$F(p,q) = B\int q(x) \log (q(x)/p(x))dx + C \int p(x) \log (p(x)/q(x))dx$$

for some constants $B,C \geq 0$, not both zero.

Outline of Proof.

Before embarking on the proof, we state and prove a lemma that will be used several times. The proof proper has been divided into four steps. The first uses the lemma and the invariance axiom, available by virtue of Theorem A, to show that f can be written as

$$f(u,v) = g(u/v)u + D(u-v) \log v$$

in terms of a constant D and a function g of one variable. The second step uses the additivity axiom and invokes the lemma twice; the conclusion is that a certain expression involving g depends on only two of the three variables it contains. Namely,

$$\begin{aligned}&[g(tu) - g(t) - g(u)]\ tu/(1-t)(1-u)\\ &\quad - [g(tv) - g(t) - g(v)]\ tv/(1-t)(1-v)\end{aligned}$$

is independent of t. The third step proceeds to a general form for f:

$$\begin{aligned}f(u,v) &= A(u - v) + B\ v \log (v/u)\\ &\quad + C\ u \log (u/v) + D(u - v) \log v.\end{aligned}$$

The last step, with the help of the positivity axiom, eliminates A and D and shows that $F(p,q)$ has the form stated in the conclusion of the theorem.

Lemma B.

Let h be a function of one real variable. Suppose the equation

$$\int h(p(x)/q(x))\, p(x)\, dx = 0$$

holds for all positive probability-density functions p and q, on some linear space, that satisfy $F(p,q) < \infty$. Then h has the form

$$h(t) = K(1/t - 1)$$

for some real number K and all $t > 0$.

Proof of Lemma.

Let q be a positive probability density on the given linear space. Let u and v be any two positive numbers with $u > 1 \geq v > 0$ and set $r = (1 - v)/(u - v)$. Then $0 \leq r < 1$ and $ru + (1 - r)v = 1$. Choose a bounded set M, a subset A, and a real number m with

$$\int_M q(x)\, dx = m > 0$$

and

$$\int_A q(x) dx = rm;$$

define

$$p(x) = \begin{cases} uq(x), & (x \in A) \\ vq(x), & (x \in M-A) \\ q(x), & (x \in \overline{M}) \end{cases},$$

where M-A is the relative complement of A in M and $\overline{M}$ is the complement of M. Then p is a positive probability density, since

$$\int p(x)\, dx = \int_A p(x)\, dx + \int_{M-A} p(x)\, dx + \int_{\overline{M}} p(x)\, dx$$

$$= \int_A u\, q(x)\, dx + \int_{M-A} v\, q(x)\, dx + \int_{\overline{M}} q(x)\, dx$$

$$= rmu + (1 - r)mv + (1-m)$$

$$= 1.$$

Moreover, $F(p,q) < \infty$, since

$$F(p,q) = \int_M f(p(x),q(x))\,dx + \int_{\bar{M}} f(q(x),q(x))\,dx,$$

where the first integral is over a bounded set and the second vanishes by Lemma A.2. Thus, p and q are positive probability densities on the given linear space and satisfy $F(p,q) < \infty$. By hypothesis, this implies

$$\int h(p(x)/q(x))\,p(x)\,dx = 0.$$

Now

$$\int h(p(x)/q(x))\,p(x)\,dx$$

$$= \int_A h(u)\,u\,q(x)\,dx + \int_{M-A} h(v)\,v\,q(x)\,dx + \int_{\bar{M}} h(1)q(x)\,dx$$

$$= rmuh(u) + (1-r)mvh(v) + (1-m)h(1).$$

Therefore

$$\frac{1-v}{u-v}\,muh(u) + \frac{u-1}{u-v}\,mvh(v) + (1-m)h(1) = 0.$$

When $v = 1$, this implies

$$h(1) = 0.$$

When $v \neq 1$, it implies

$$h(u)\,u/(1-u) = h(v)v/(1-v),$$

and then both sides equal some constant K, since one side is independent of u and the other is independent of v. Consequently,

$$h(u) = K(1/u - 1),$$
$$h(v) = K(1/v - 1)$$

whenever $u > 1$ and $0 < v \leq 1$. Therefore

$$h(t) = K(1/t - 1)$$

for every $t > 0$.

Proof of Theorem.

Step 1. Theorem A implies that the invariance axiom holds for p and q such that $F(p,q) < \infty$. To get information about $f(u,v)$ for some particular real numbers u and v, we will apply the axiom to density functions that assume constant values on certain intervals.

Let $t > 0$. Multiplication by t is a nonsingular linear transformation T in one dimension: $Tx = tx$. The axiom in this case implies

$$\int f(p(x),\ q(x))\ dx = \int f(tp(tx),\ tq(tx))\ dx$$

$$= \int f(tp(x),\ tq(x))(1/t)\ dx.$$

where the second equality results from a change of variables: $(1/t)x$ for x.

Let $u > 0$ and $s > 0$. Define $p(x) = u$ and $q(x) = 1$ when $0 \le x \le a$, where a is chosen so that $au < 1$ and $a < 1$. Off the interval $[0,a]$ define p and q in any way that makes them positive probability-density functions on the real numbers and makes $F(p,q)$ finite. Define

$$p'(x) = \begin{cases} p(x), & x < 0 \\ su, & 0 \le x \le a/s \\ p(a + x - a/s), & a/s < x \end{cases}$$

$$q'(x) = \begin{cases} q(x), & x < 0 \\ s, & 0 \le x \le a/s \\ q(a + x - a/s), & a/s < x \end{cases}$$

(cf. figure 1 and the proof of Theorem A).

Then $\int p'(x)\ dx = \int q'(x)\ dx = 1$, since the integrals of p' and q' over the subintervals $(-\infty, 0)$, $[0, a/s]$, and $(a/s, \infty)$ are respectively equal to the integrals of p and q over $(-\infty,0)$, $[0,a]$, and (a, ∞). Thus p' and q' are positive probability-density functions. Moreover, $F(p',q') < \infty$. We have

$$\int f(p(x),\ q(x))\ dx - \int f(p'(x),\ q'(x))\ dx$$

$$= af(u,1) - af(su,s)/s,$$

since the parts of the first integral outside $[0,a]$ cancel the parts of the second outside $[0,a/s]$. Similarly,

$$\int f(tp(x), tq(x))\,(1/t)\,dx - \int f(tp'(x), tq'(x))\,(1/t)\,dx$$

$$= af(tu,t)/t - af(stu,st)/st.$$

Therefore, by the invariance axiom,

$$f(u,1) - f(su,s)/s$$

$$= f(tu,t)/t - f(stu,st)/st.$$

For any given fixed positive value of u, define

$$k(s) = f(u,1)/u - f(su,s)/su\ ;$$

it follows that

$$k(s) + k(t) = k(st)\ ,$$

for arbitrary $s,t > 0$, and k is therefore of the form

$$k(s) = h \log s,$$

where h is a real number. We write h(u) for h, since h may depend on the given value of u. Then, for any positive numbers u and v, we have

$$f(u,1)/u - f(su,s)/su = h(u) \log s.$$

Define

$$g(u) = f(u,1)/u\ ;$$

then, substituting u/v for u and v for s, we find that f has the general form

$$f(u,v) = g(u/v)u - h(u/v)\,u \log v$$

for arbitrary $u,v > 0$.

Now the invariance axiom implies

$$\int [g(p(x)/q(x)) - h(p(x)/q(x)) \log q(x)]\,p(x)\,dx$$

$$= \int [g(tp(x)/tq(x)) - h(tp(x)/tq(x)) \log tq(x)]\,tp(x)(1/t)\,dx,$$

which simplifies to

$$(\log t)\int h(p(x)/q(x))\, p(x)\, dx = 0.$$

Applying the lemma results in

$$h(t) = D(1/t - 1)$$

for some real number D. It follows that

$$f(u,v) = g(u/v)\, u + D(u - v) \log v.$$

Step 2. Let p', p", q', and q" be positive probability-density functions such that $F(p',q') < \infty$, $F(p",q") < \infty$. The additivity axiom states that

$$\int f(p'(x')p''(x''),\ q'(x')q''(x''))\, dx'dx''$$

$$= \int f(p'(x'),\ q'(x'))\, dx' + \int f(p''(x''),\ q''(x''))\, dx''.$$

Substituting the expression for f from step 1 gives

$$\int [g(p'(x')p''(x'')/q'(x')q''(x''))\, p'(x')\, p''(x'')$$

$$+\ D(p'(x')\, p''(x'') - q'(x')\, q''(x''))\log q'(x')q''(x'')]\, dx'dx''$$

$$= \int [g(p'(x')/q'(x'))\, p'(x') + D(p'(x') - q'(x'))\log q'(x')]\, dx'$$

$$+ \int [g(p''(x'')/q''(x''))\, p''(x'') + D(p''(x'') - q''(x''))\log q''(x'')]\, dx''.$$

This simplifies to

$$\int [g(p'(x')p''(x'')/q'(x')q''(x''))$$

$$-\ g(p'(x')/q'(x')) - g(p''(x'')/q''(x''))]\, p'(x')\, p''(x'')\, dx'dx'' = 0.$$

For any given fixed p',q' as above, define a new h by setting

$$h(t) = \int [g(tp'(x')/q'(x'))$$

$$- g(p'(x')/q'(x')) - g(t)]\ p'(x')\ dx'.$$

Then

$$\int h(p''(x'')/q''(x''))\ p''(x'')\ dx'' = 0.$$

By the lemma, $h(t) = K(1/t - 1)$ for some real number K (which may depend on the given p' and q'). Whatever the value of K, we have

$$h(1) = 0,$$

$$uh(u)/(1 - u) = vh(v)/(1 - v)$$

when $u \neq 1$ and $v \neq 1$. The first of these equations implies

$$g(1) = 0.$$

The second, when expanded, becomes

$$\int\Big\{[g(up'(x')/q'(x')) - g(p'(x')/q'(x')) - g(u)]\ u/(1-u)$$

$$- [g(vp'(x')/q'(x')) - g(p'(x')/q'(x')) - g(v)]\ v/(1-v)\Big\}\ p'(x')\ dx' = 0.$$

This holds whenever p' and q' are positive probability-density functions with $F(p',q') < \infty$.

Now, for any given fixed values of u and v (positive, not 1), define a new h:

$$h(t) = [g(tu) - g(t) - g(u)]\ u/(1-u)$$

$$- [g(tv) - g(t) - g(v)]\ v/(1-v).$$

Then

$$\int h(p'(x')/q'(x'))\ p'(x')\ dx' = 0.$$

By the lemma, $h(t) = K(1/t - 1)$ for some real number K. We write $K(u,v)$ for K, since K may depend on the given values of u and v. Then

$$[g(tu) - g(t) - g(u)]\ tu/(1-t)(1-u)$$
$$- [g(tv) - g(t) - g(v)]\ tv/(1-t)(1-v)$$
$$= K(u,v)$$

for any positive t, u, and v different from 1.

Step 3. Define

$$k(u,v) = [g(uv) - g(u) - g(v)]\ uv/(1-u)(1-v).$$

The result of step 2 is

$$k(t,u) - k(t,v) = K(u,v).$$

Therefore, for $t = v$ and $t = u$, respectively,

$$k(v,u) - k(v,v) = K(u,v) = k(u,u) - k(u,v).$$

But $k(u,v) = k(v,u)$; hence

$$2k(u,v) = k(u,u) + k(v,v).$$

Define a function

$$a(u) = k(u,u)(1 - u)/2.$$

Then

$$k(u,v) = a(u)/(1 - u) + a(v)/(1 - v).$$

By the definition of $k(u,v)$,

$$uv\ [g(uv) - g(u) - g(v)] = a(u)(1 - v) + a(v)(1 - u). \tag{4}$$

(This holds even when $u = 1$ or $v = 1$, since $g(1) = 0$.) Substituting t for u and uv for v leads to

$$tuv\ [g(tuv) - g(t) - g(uv)] = a(t)(1-uv) + a(uv)(1-t);$$

this, with the previous equation, gives

$$tuv\ [g(tuv) - g(t) - g(u) - g(v)]$$
$$= a(t)(1 - uv) + a(uv)(1-t) + t\ a(u)(1 - v) + t\ a(v)(1 - u).$$

Consequently,

$$tuv\,[g(tuv) - g(t) - g(u) - g(v)]$$

$$- a(t)(1 - uv) - a(u)(1 - tv) - a(v)(1 - tu)$$

$$= a(uv)(1 - t) - a(u)(1 - t) - a(v)(1 - t)$$

$$= a(tu)(1 - v) - a(t)(1 - v) - a(u)(1 - v),$$

where the second equality is due to the symmetry of the left-hand side in t, u, and v. It follows that

$$\frac{a(tu) - a(t) - a(u)}{(1 - t)(1 - u)} = \frac{a(vu) - a(u) - a(v)}{(1 - u)(1 - v)} = \frac{a(tv) - a(t) - a(v)}{(1 - t)(1 - v)} ,$$

where the second equality is again due to symmetry. The left-hand side is thus independent of t and u; it is therefore a constant, which with foresight we denote A/2:

$$\frac{a(tu) - a(t) - a(u)}{(1 - t)(1 - u)} = A/2$$

Hence

$$a(tu) - a(t) - a(u) = (A/2)\,[-(1 - tu) + (1 - t) + (1 - u)].$$

This can be written as

$$b(t) + b(u) = b(tu) ,$$

where the function b is defined by

$$b(t) = -a(t) - (A/2)(1 - t).$$

It follows that

$$b(t) = B \log t$$

for some constant B. Thus (5)

$$a(t) = -(A/2)(1 - t) - B \log t.$$

Define a function

$$c(t) = g(t) + A(1/t - 1) + B(\log t)/t.$$

We have

$$uv\,[c(uv) - c(u) - c(v)]$$

$$= uv\,[g(uv) - g(u) - g(v)]$$

$$+ Auv\,[(1/uv - 1) - (1/u - 1) - (1/v - 1)]$$

$$+ Buv\,[(\log uv)/uv - (\log u)/u - (\log v)/v]$$

$$= a(u)(1 - v) + a(v)(1 - u)$$

$$+ A(1 - u)(1 - v)$$

$$+ B(\log u)(1 - v) + B(\log v)(1 - u)$$

$$= 0$$

with the help of (4) and (5); thus $c(uv) - c(u) - c(v) = 0$. Therefore

$$c(t) = C \log t$$

for some constant C. This yields

$$g(t) = -A(1/t - 1) - B(\log t)/t + C \log t.$$

Finally, by the result of Step 1,

$$f(u,v) = A(u-v) + B\,v \log (v/u)$$

$$+ C\,u \log (u/v) + D(u-v)\log v.$$

Step 4. By the result of step 3,

$$F(p,q) = A\int (p(x) - q(x))dx + B\int q(x) \log (q(x)/p(x))\, dx$$

$$+ C\int p(x) \log (p(x)/q(x)) + D\int (p(x) - q(x)) \log q(x)\, dx.$$

The first term on the right is zero and may be deleted, since $\int p(x)\, dx = 1 = \int q(x)\, dx$. The coefficients B, C, and D of the remaining terms are restricted by the positivity axiom. In fact, D must be zero.

To show this, we start by choosing functions q and r that satisfy three requirements:

(1) $q + tr$ is a positive probability density for all t in some neighborhood of 0,

(2) $\int r(x) \log q(x)\, dx \neq 0$,

(3) differentiation under the integral sign is permissible in computing $dF(q + tr, q)/dt$.

The first requirement implies that q is a positive probability density and that $\int r(x)\,dx = 0$. The differentiation gives

$$\frac{dF(q + tr, q)}{dt} = - B \int \frac{q(x)r(x)}{q(x) + tr(x)}\,dx$$

$$+ C\int r(x) \left[1 - \log \frac{q(x)}{q(x) + tr(x)}\right] dx$$

$$+ D\int r(x) \log q(x)\,dx \quad .$$

When t = 0, this reduces to

$$(C - B)\int r(x)\,dx + D\int r(x) \log q(x)\,dx$$

The first of the three requirements implies that the first term vanishes. The second requirement implies that the second term is nonzero unless D = 0.

Suppose $D \neq 0$. Then, when t = 0, we have $F(q + tr,q) = F(q,q) = 0$, but $dF(q + tr,q)/dt \neq 0$. It follows that $F(q + tr,q)$ assumes both positive and negative values as t varies in some neighborhood of 0. This contradicts the positivity axiom. Therefore D = 0. Thus

$$F(p,q) = B\int q(x) \log(q(x)/p(x))\,dx$$

$$+ C\int p(x) \log (p(x)/q(x))\,dx$$

Both integrals are nonnegative and vanish only if p = q. By suitable choice of p and q, either integral may be made arbitrarily large in comparison with the other. Therefore, the positivity axiom requires that B and C be both nonnegative and at least one positive.

Corollary.

Theorem B remains true if, in the hypotheses, the positivity axiom is replaced with the following axiom.

Semiboundedness. $F(p,q) \geq F(p,p)$ with equality only if p = q.

Proof.

Assume the modified hypotheses. By Lemma A.2, $F(p,p) = 0$; consequently, the semiboundedness axiom implies the additivity axiom. The unmodified hypotheses of Thoerem B are thus satisfied. Therefore the conclusion holds.

IV. DISCUSSION

The theorem just proved shows that in computing the functional $F(p,q) = \int f(p(x),q(x))dx$, we may take

$$f(u,v) = B\, v \log (v/u) + C\, u \log (u/v) \tag{6}$$

when $u > 0$ and $v > 0$. The possible extension of F to densities p and q that may assume the value 0 was deferred to this section. The expression for $f(u,v)$ prepares one to expect infinities to crop up when the arguments of f become 0.

As a matter of fact, the functional F, though not the function f, may become infinite even when we keep the restrictions $p(x) > 0$, $q(x) > 0$. Kullback [4, pp. 6, 10] points out the problem and gives some examples. Here is one more. Let

$$p(x) = (2\pi)^{-\frac{1}{2}} \exp(-x^2/2) \quad ,$$

$$q(x) = \pi^{-1} (1 + x^2)^{-1} \quad .$$

Then one directed divergence is infinite and the other is finite:

$$\int q(x) \log(q(x)/p(x))\, dx = \infty \quad ,$$

$$\int p(x) \log(p(x)/q(x))\, dx < \infty \quad .$$

Setting

$$p'(x) = q'(-x) = \begin{cases} p(x), & x > 0 \\ q(x), & x \le 0 \end{cases}$$

gives densities p' and q' whose directed divergences are both infinite.

Granted, the integrals may diverge, but at worst they evaluate unambiguously to $+\infty$; the indeterminate form $\infty - \infty$ does not occur. Namely the negative part of the integral -- the integral over the set where the integrand is negative -- is finite. Consider $\int p(x) \log(p(x)/q(x))\, dx$, for example. We have $p(x) \log(p(x)/q(x)) \ge (-1/e)q(x)$, since the minimum of $t \log t$ for $t > 0$ is the value $-1/e$ at the point $t = 1/e$. Therefore, the negative part of the integral is not less than $-1/e$.

The axioms thus remain meaningful for infinite values of F. For instance, the equation $F(p',q') = F(p,q)$ in the invariance axiom implies that if either of the quantities $F(p',q')$ and $F(p,q)$ is infinite, then so is the other. With f defined as in (6), one can check that the axioms not merely remain meaningful, they are actually satisfied for all positive probability densities p and q.

We now come to the question of probability densities p and q that may take the value 0. Is it possible to define $f(0,0)$, $f(u,0)$, and $f(0,v)$ for positive real numbers u and v so that the axioms still hold? The answer is no if we continue to insist that f be a real-valued function. It can be shown that if $C > 0$ in (6), then the axioms are inconsistent with having $f(u,0) < \infty$ for $u > 0$. Likewise, if $B > 0$ in (6), then the axioms rule out $f(0,v) < \infty$ for $v > 0$. If we permit f to take an infinite value when one or both of its arguments are zero, we may adopt the convention

$$u \log (u/0) = \infty$$

when $u > 0$. This convention is natural, since $u \log(u/t)$ tends to ∞ as t approaches 0. Similarly, the conventions

$$0 \log (0/u) = 0$$

$$0 \log (0/0) = 0$$

are the natural ones.

ACKNOWLEDGEMENTS

These results were obtained in connection with joint work on prior probabilities and entropy maximization with J. Shore, whom the author thanks for continuing helpful discussion and for critical readings of earlier versions of this paper. The author thanks W. S. Ament for discussions and J. Morris for a thorough review.

REFERENCES

1 R. L. Kashyap, "Prior Probability and Uncertainty," IEEE Trans. Information Theory, IT-17, 641-650 (November 1971).

2 E. T. Jaynes, "Prior probabilities", IEEE Trans. Syst. Sci. Cybern., SSC-4, 227-241 (September 1968).

3 R. Johnson, "Comments on 'Prior Probability and Uncertainty'", submitted for publication.

4 S. Kullback, Information Theory and Statistics, Dover, New York, 1957.

5. Pl. Kannappan, "On Shannon's entropy, directed divergence, and inaccuracy", Z. Wahrscheinlichkeitstheorie u. Verw. Geb. 22, 95-100 (1972).

6. Pl. Kannappan, "On directed divergence and inaccuracy", Z. Wahrscheinlichkeitstheorie u. Verw. Geb. 25, 49-55 (1972).

7. J. Aczél and Z. Daróczy, On Measures of Information and Their Characterizations, Academic Press, New York. 1975.